ABSTRACT OF THE DISCLOSURE

A method for computing a diversity measure H(m) for combinatorial structures involves identifying all M possible substructures having m elements from among the n elements of the combinatorial structure. The number of the substructures that are similar to each such substructure is determined, and the frequency of each distinct substructure is calculated using the number of similar substructures and the total number of substructures M. The method uses the frequency of each distinct substructure to compute an entropy corresponding to m. By the same process described above, and entropy corresponding to m+1 is computed. The entropy corresponding to m+1 is subtracted from the entropy corresponding to m to produce the diversity measure H(m). In the preferred embodiment, similar substructures are determined by being identical or isomorphic. In an alternative embodiment, a distance function is used to compute a distance between two substructures, and only if the distance is less than a predetermined threshold are the two substructures determined to be similar. In the preferred embodiment, the entropy is computed by summing the frequency of each distinct substructure multiplied by the logarithm of the frequency of each distinct substructure. In an alternative embodiment, the entropy is computed by summing the frequency of each distinct substructure by the logarithm of the quotient of the frequency divided by an expected frequency of the distinct substructure. Generalized graphs such as can be used to model the Web are combinatorial structures suitable for use with the methods according to the present invention.